*Citation for published version:*
Agarwal, R & Gaule, P 2020, 'Invisible Geniuses: Could the Knowledge Frontier Advance Faster?', *American Economic Review: Insights*, vol. 2, no. 4, pp. 409-424. https://doi.org/10.1257/aeri.20190457

*DOI:*
10.1257/aeri.20190457

*Publication date:*
2020

*Document Version*
Peer reviewed version

Link to publication

This is the author accepted manuscript of an article published in final form in American Economic Review Insights, Vol. 2, no. 4 (December 2020), pp. 409-424. https:/dx.doi.org/10.1257/aeri.20190457

**University of Bath**

## Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

# Invisible Geniuses: Could the Knowledge Frontier Advance Faster?

Ruchir Agarwal and Patrick Gaule*

January 28, 2020

**Abstract**

A better understanding of the determinants of idea/knowledge production remains critical for long-run growth. Towards this end, this paper establishes two results using data from the International Mathematical Olympiad (IMO). First, individuals who excelled in teenage years are especially capable of advancing the knowledge frontier. Second, such talented individuals born in poorer countries are systematically less likely to engage in knowledge production. IMO participants from low-income countries produce 34% fewer publications and 56% fewer cites than equally talented rich-country counterparts. Policies to encourage talented youth to pursue scientific careers–especially those from poorer countries–could advance the knowledge frontier faster.

*Agarwal: International Monetary Fund, RAgarwal@imf.org. Gaule: University of Bath, P.Gaule@bath.ac.uk.
Acknowledgements: We appreciate helpful comments from Dan Cao, Christian Catalini, Evan Chen, Annamaria Conti, Tom Cunningham, Kirk Doran, Ehsan Ebrahimy, Tarhan Feyzioglu, Christian Fons-Rosen, Ina Ganguli, Jon Gruber, Niels-Jakob Hansen, Scott Kominers, Alessandro Iara, Artjom Ivlevs, Xavier Jaravel, Vitalijs Jascisens, Ben Jones, Stepan Jurajda, Jin Li, Megan MacGarvie, Mark McCabe, Nikolas Mittag, Abishek Nagaraj, Alex Oettl, Chris Peterson, Pian Shu, Geoff Smith, Sarah Smith, John von Reenen, Valentina Tartari, Otto Toivanen, Fabian Waldinger, Heidi Williams as well as from seminar and conference participants at the NBER Summer Institute, Barcelona Summer Forum, MIT Sloan School, Georgia Tech, the Higher School of Economics, the University of Bristol, Copenhagen Business School, the University of Leicester, the University of Massachusetts and ZEW Mannheim. We are grateful to Suzie Maine and Deborah Novakova for excellent work in proofreading and copy-editing the paper. All errors are our own. The views expressed therein are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. Gaule acknowledges financial support from the Czech Science Foundation (GACR grant no 16-05082S).

In many growth models, economic growth arises from people creating knowledge, and the long-run growth rate is the product of two terms: the number of researchers and their research productivity. Bloom et al. (2017) present a wide range of evidence from various industries showing that the productivity of researchers is declining sharply-thereby making new knowledge or groundbreaking ideas harder and harder to find. Thus, a better understanding of the determinants of idea/knowledge production remains critical. In this context, this paper asks two questions: (1) Could encouraging talented youth to pursue research careers play a critical role in knowledge creation? (2) If so, how does the country in which a talented youth is born influence the quantity of knowledge she produces over her lifetime?

We focus on these questions because they are central to understanding both the process of knowledge creation and whether the world is utilizing the pool of talented youth optimally to advance the knowledge frontier.[1] Answering these questions raises three major empirical challenges: (1) how to measure talent; (2) how to make this measurement of talent comparable across multiple countries and over time; and (3) how to construct a broad sample of talented youth without selecting on eventual lifetime success in knowledge production. To address these challenges, we focus on knowledge production in mathematics and use a unique institutional feature of this discipline: the International Mathematical Olympiads (IMO), a prominent worldwide competition for high-school students. This setting allows us to measure talent in teenage years (as proxied by IMO scores) as well as to conduct direct comparisons of talent in teenage years across countries. By connecting multiple sources, we are able to build an original database covering the education history and publications of the population of IMO participants participating across 20 years of the

[1] There has been little systematic study of these questions with some notable exceptions. Aghion et al. (2017) find a significant but relatively weak correlation between visiospatial IQ (from an army entrance exam) and the propensity to become an inventor in Finland. Bell et al. (2019) also report a correlation between 3rd grade math scores and the propensity to become a patent inventor in the U.S. There is also a psychology literature investigating the nexus between intelligence, creativity and scientific achievement. For instance, Cox (1926) estimates IQ scores for 300 ‘geniuses’ who made outstanding contributions to science.

competition (1981-2000; n=4,710). In this paper, we use the word talent to refer to an individual's problem-solving capacity in their teenage years. This could be a product of innate ability, practice, or both. For the purposes of this paper, we are agnostic about the relative importance of innate ability and environment in determining teenage talent.

We first document a salient positive correlation between the points scored at the IMO–and particularly the points scored on the most difficult problems–and subsequent mathematics knowledge production. Strong IMO performers are significantly more likely to produce frontier research in mathematics than either lesser IMO performers or mathematicians who did not participate in the IMO. The conditional probability that an IMO gold medalist will become a Fields medalist is *fifty* times larger than the corresponding probability for a PhD graduate from a top 10 mathematics program.

Next, we investigate the role of country of origin on career outcomes and knowledge production of IMO participants controlling for IMO score. We find that there is a developing country penalty throughout the talent distribution in our sample. That is, compared to their counterparts from high-income countries who obtained the same score in the IMOs, participants born in low- or middle-income countries produce considerably less knowledge over their lifetime. A participant from a low-income country produces 34% fewer mathematics publications and receives 56% fewer mathematics citations than an equally talented participant from a high-income country. A back-of-the-envelope calculation suggest the knowledge production (from IMO participants) could be 10% higher in terms of publications and 17% higher in terms of cites.[2]

It is interesting to consider whether differences in socioeconomic background within the U.S. are more important or less important than international differences among countries in generating differences in innovation outcomes. The magnitudes of the low-income country penalty we report

[2]The details of the calculation and the associated limitations are discussed in section IV.

are an order of magnitude lower than those reported in Bell et al. (2019) who find children from high-income (top 1%) families are 10 times as likely to become inventors as those from below median income families. However, it is important to keep in mind that while they look at all children born in the U.S., we only consider individuals who have achieved excellence in their late teens. Such individuals have already overcome many barriers from a less favorable environment in accumulating human capital. In addition, we are looking at individuals who are the tail of the talent distribution and one would expect those to be the most likely to be able to overcome barriers. For these reasons, one should think of our estimates as a lower bound on the impact of coming from a poor country–which does not fully account for the extensive margin of barriers to opportunity.

This work builds upon the macroeconomic literature on talent allocation and the microeconomic literature on the origin of knowledge producers.[3] Baumol (1990) and Murphy, Shleifer and Vishny (1991) emphasize the allocation of talent across different sectors of the economy as being key for economic growth. More recently, Hsieh et al. (2019) attribute part of aggregate wage growth in the U.S. to the integration of talented women and blacks into the U.S. labor market. Using PISA data, Ugarov (2019) shows that in developing countries academic skills are less predictive of occupational choice that in developed countries, possibly due to occupational barriers. Most relevant for us, recent empirical literature investigates how children's socioeconomic and geographic backgrounds influence their likelihood of becoming a patent inventor in the U.S. (Akcigit, Grigsby and Nicholas 2017; Celik 2017; Bell et al. 2019) and in Finland (Aghion et al. 2017). A consistent finding is that children of low-income parents are much less likely to become inventors than their higher-income counterparts. Bell et al. (2019) also report considerable differences between states of birth in the likelihood of becoming an inventor. For instance, they find that children born in Massachusetts are five times more likely than children born in Alabama to become

[3] We also build on the literature on the role of place in knowledge production (Kahn and MacGarvie 2016), and on the determinants of high math achievement (Andreescu et al. 2008; Ellison and Swanson 2010, 2016, 2018).

inventors. Our results echo these differences at the international level. Besides the cross-country dimension, a distinguishing feature of our data is that we have a sample of individuals in the very right tail of the ability distribution; and we document that, even there, different backgrounds result in substantial differences in knowledge produced. The evidence we present on the importance of talent in the production of frontier knowledge also sheds light on the costs associated with the misallocation of talent away from knowledge producing careers.

More generally, our results also relate to the study of the determinants of the rate of knowledge production. The endogeneous growth literature has studied the size of the knowledge production sector (see e.g. Jones 2002, Freeman and Van Reenen 2009, Bloom et al. 2017) but has given less attention to its composition. Similarly, the literature on the economics of science has typically focused on how institutions and incentives affect the productivity of existing researchers rather than on who becomes a scientist or knowledge producers in the first place.[4] Our study suggests that the selection of talented individuals into knowledge production may be important for the rate of scientific progress. Lastly, our paper also builds on the pull vs. push incentives literature (Maurer 2006; Kremer and Williams 2010 ; Fu, Lu and Lu 2012; Williams 2012). While pull incentives can be potentially important for stimulating scientific discoveries, our study suggests that push programs targeted at talented youth could be an effective complementary tool for the advancement of the knowledge frontier.

The paper proceeds as follows. Section I describes the International Mathematical Olympiad. Section II presents the data. The results on the link between IMO success and long-term achievements appear in section III and those on cross-country comparisons in section IV. Section V con-

[4]For instance, Borjas and Doran (2012) study the productivity of U.S. mathematicians following a large influx of Russian mathematicians into the U.S. Other studies on the determinants of scientific productivity among established scientists include Azoulay, Graff Zivin and Wang 2010; Waldinger 2010, 2011; Azoulay, Graff-Zivin and Manso 2011; Jacob and Lefgren 2011; Ganguli 2017; Iara, Schwarz and Waldinger 2018. For a more general survey on the economics of science, see Stephan 2012.

cludes.

# I The International Mathematical Olympiads

The International Mathematical Olympiad (IMO) is a prominent mathematics competition held annually since 1959. Participants travel to the location of the IMO (a different city every year) together as part of a national team. Expenses for the participants, including travel and accommodation are paid for the by the organization hosting the IMO in that year. The competition is aimed at high school students, and requires that participants be younger than 20 years of age and not enrolled at a tertiary education institution. Initially, only Eastern European countries sent participants, but over time participation expanded to include over 100 countries.[5]

The IMO participants are selected by their national federation (up to six per country), normally on the basis of regional and national competitions. As countries vary in size and in popularity of mathematics competitions, becoming an IMO participant is more difficult in some countries than others. In the empirical exercise, we will compare IMO participants controlling for IMO performance. This should reduce the effect of this differential selectivity. Some participants compete in several successive years but the majority of participants only compete once. They solve a total of six problems drawn from geometry, number theory, algebra and combinatorics. Each problem is worth seven points and participants can score up to 42 points. Medals (gold, silver or bronze) are awarded based solely on the sum of points collected across problems. Slightly fewer than half of the participants receive medals, which are gold, silver or bronze.

The same process is used every year to generate a set of novel problems of appropriate diffi-

[5] The United Kingdom and France joined in 1967, the U.S. joined in 1974 and China in 1985; they have participated regularly since. The only countries with population above 20 million that have never participated are Ethiopia, Sudan and the Democratic Republic of Congo.

culty, and to grade them. There is a sense among observers, however, that the IMO has become somewhat more difficult over time.

# II Data

Multiple sources of data were combined to create the original data for this paper. We extracted data on all IMO participants from the official IMO website (http://www.imo-official.org) and then selected those who participated between 1981 and 2000, inclusive. Some participants compete in multiple years, in which case we kept only the last participating year. We ended up with a list of 4,710 individuals.

We then constructed long-term performance outcomes in mathematics for these individuals using PhD and bibliometric data. For PhD theses, we relied on the Mathematics Genealogy Project, a volunteer effort aimed 'to compile information on all mathematicians in the world'.[6] It has achieved broad coverage, with information on more than 200,000 mathematicians. For each graduating student, it lists university, advisor name, graduation year and dissertation topic. For bibliometric data, we used MathSciNet data which is produced by Mathematical Reviews under the auspices of the American Mathematical Society. While the underlying publication data is richer, our outcomes are based on total publications and cites by author as computed by MathSciNet (and reflecting the manual author disambiguation by the publishers of Mathematical Reviews). Both of these databases have been used in prior research (Borjas and Doran 2012, 2015a, 2015b; Agrawal, Goldfarb and Teodoridis 2016). We complemented the publication data by collecting a list of speakers at the International Congress of Mathematicians (ICM) and tagging IMO participants

[6]One might worry that the coverage of the Mathematics Genealogy Project might be worse for developing countries, which would be problematic for cross-country comparisons. However, we have encountered only a handful of individuals with math publications (or with a faculty appointment in mathematics) that were not listed in the genealogy project, and they were mostly from developed countries.

who were speakers at the ICM congress. Being invited to speak at the ICM congress is a mark of honor for mathematicians and we use it a measure of community recognition independent of bibliometrics. Similarly, we tag IMO participants who have received the Fields medal.

While most of this paper focuses on those bibliometric and PhD data, we also manually searched the names of IMO participants online to construct current employment measures. Given that this part of the data collection is particularly time-intensive, this information was only collected for IMO medalists (2,272 people out of the 4,710 participants). We coded whether the person is employed in mathematics academia, employed in academia outside mathematics, employed in industry, or does not have an online profile.

Our final database covers the population of IMO participants who participated between 1981 and 2000 (4,710 people).[7] Besides information on IMO participation (year, country, points scored, type of medal), we know whether the person holds a PhD in mathematics, and, if so, what year and from what school it was earned, their mathematics publications and cites counts until 2015 and whether that person has been a speaker at the ICM Congress or a Fields Medalists. For IMO participants who hold a PhD in mathematics, we are interested in whether they have graduated from an elite school; we proxied this by graduating from one of the top ten schools in the Shanghai 2010 mathematics rankings.

Descriptive statistics on our sample are available in appendix table A1. Around 8% of IMO participants earn a gold medal, while 16% have a silver medal and 24% have a bronze medal; a further 10% have an honourable mention. Around 22% of IMO participants hold a PhD in mathematics; of those around a third a PhD in mathematics from a top 10 school. One percent of IMO participants became ICM speakers, and 0.2% became Fields medalists. Collectively, the

[7] Our focus on these earlier cohorts ensures that we have at least 15 years of post-IMO participation for which measure outcomes. Ninety percent of IMO participants who ended up earning a PhD in mathematics do so within 13 years of their last IMO participation, with no systematic variation across country income groups.

IMO participants in our sample produced more than 15,000 publications and received more than 160,000 cites.

We proxied an individual's country of origin by the country the individual represented at the IMO. Around half of the participants were from high-income countries (as per the 2000 World Bank classification), 18% from upper middle-income countries, 23% from lower middle-income countries, and 11% from low-income countries.

Finally, we produced an ancillary dataset that has all the PhD graduates (irrespective of IMO participation) listed in the Math Genealogy Project who graduated between 1990 and 2010 (n=89,086). For those, we know the school and year they graduated, how many math publications and cites they produced (from MathSciNet) and whether they have been ICM speakers or Fields Medalists.

# III How do IMO scores relate to long-term performance?

In this section, we investigate whether teenage talent – as proxied by IMO scores – is correlated with long term performance in the field of mathematics. On one hand, it is natural to expect performance at the IMO to be positively correlated with becoming a professional mathematician, mathematical knowledge production, and outstanding achievements in mathematics. After all, both IMO problems and research in mathematics are two activities that involve problem solving in the field of mathematics. Moreover, there is ample anecdotal evidence of distinguished mathematicians having won IMO medals.

However, the link between IMO score and long-term performance is less obvious than it seems. Mathematical research encompasses activities other than problem solving - such as conceptualization and hypothesis generation. Moreover, the IMOs problems are quite different from research problems, in that they are known to have a solution and they have to be solved in a very short time frame, without access to literature or the help of other mathematicians. Even if talent is indeed important for the production of knowledge, IMO scores may not be a good measure of talent. If some measure of luck or extraeneous factors affect IMO scores, this will introduce classical measurement error when considering IMO scores as a measure of talent.[8]

We begin with some graphical evidence: Figure 1 plots the mean achievements of IMO participants by the number of points they scored at the IMO, with a linear fit superimposed.[9] Six achievements are considered: obtaining a PhD degree in mathematics, obtaining a PhD in mathematics from a top 10 school[10], the number of mathematics publications (in logs), the number of mathematics cites (in logs), being a speaker at the International Congress of Mathematicians, and earning a Fields medal.

(insert Figure 1 about here)

The graphs in the first two rows of Figure 1 all display a clear positive gradient: IMO participants with higher IMO scores are more likely to have a PhD in mathematics or a PhD in mathematics from a top school; and they produce more mathematical knowledge measured in terms of publications and cites. The two graphs in the bottom rows measure exceptional achievements in mathematics. By definition, these are rare outcomes and the graphs are less smooth, but the broad patterns are similar.

We also investigate the relationship between points scored at the IMO and subsequent achievements in a regression format at the individual level. We regress each achievement on points scored

[8]We also have to allow for the possibility that IMO scores may be correlated with performance in mathematics research even if IMO scores are uninformative about talent. That scenario could occur if IMO performance had a causal effect on performance. In the appendix, we present the results of a regression discontinuity comparing individuals who narrowly made versus narrowly missed medal thresholds.

[9]This figure, as others in the paper, benefits from the 'plotplain' Stata scheme developed by Daniel Bischof (see Bischof 2017).

[10]We define top 10 schools based upon the Shanghai 2010 mathematics rankings. These include Princeton, Berkeley, Harvard, Stanford, Cambridge, Paris 6, Oxford, MIT, Paris 11 and UCLA.

at the IMO, controlling for cohort and country of origin:

$$Y_{it} = \beta IMOscore_{it} + \delta X_{it} + \varepsilon_{it} \tag{1}$$

where $i$ indexes IMO participants and $t$ indexes Olympiad years. $Y_{it}$ is one of the six outcomes variables as previously defined, $IMOscore_{it}$ is the number of points scored at the IMO; $X_{it}$ includes cohort (Olympiad year) fixed effects and country of origin fixed effects.

Results (see appendix table A2) suggests that each additional point scored at the IMO (out of a total possible score of 42 - the sample mean is 16 and standard deviation 11) is associated with a 1 percentage point increase in likelihood of obtaining a Ph.D., a 2.6 percent increase in publications, a 4.3 percent increase in citations, a 0.1 percentage point increase in the likelihood of becoming an ICM speaker, and a 0.03 percentage point increase in the likelihood of becoming a Fields medalist. Furthermore, the number of points score in more difficult problems is more predictive of future achievements than the number of points scored in the easier problems (appendix Table A3).

We also compare IMO medalists to mathematicians who never went to the IMO. To do this, we constructed a sample with all PhD students obtaining a PhD in mathematics between 1990 and 2010 (n=89,068). We also constructed the subsample of PhD students graduating from top 10 schools (n=9,049). We then constructed the sample of IMOs bronze and silver medalists with a mathematics PhD (n=520) and that of gold medalists with a mathematics PhD (n=145). We plot (see Figure 2) the average number of publications, the average number of citations, the share becoming ICM speakers and the share becoming Fields medalists across the four groups.

(insert Figure 2 about here)

For each outcome, we observe the same pattern: the medalists (especially the gold medalists) outperform both other PhD graduates and the PhD graduates from top schools. While the IMO medalists produce more papers and receive more cites than other graduates, we observe a much larger difference for exceptional achievements such as being invited to the ICM Congress and receiving the Fields medal. The conditional probability that an IMO gold medalist will become a Fields medalist is fifty times larger than the corresponding probability for a PhD graduate from a top 10 mathematics program. This evidence suggests that talent may be even more important for exceptional research achievements than more routine knowledge production.

# IV Does the link between IMO score and performance depend on country of origin?

The previous section establishes that performance at the IMOs is strongly correlated with getting a PhD in mathematics and mathematics knowledge produced. We now proceed to study how this link varies according to the country of origin of IMO participants. Because we have relatively few participants for any country, we group countries in terms of income group levels (according to the 2000 World Bank classification) as a broad proxy of differences in opportunities and environment across countries. We consider specifically how IMO participants from low- and middle-income countries – about half of our sample – perform in the long-run compared to observationally equivalent participants from high-income countries. While our regressions explicitly control for IMO scores, it is worth noting that participants from developing countries do not score lower at the IMO than participants from developed countries (cf Table 1).[11]

(insert Table 1 and Figure 3 about here)

We begin by exploring graphically the link between points scored at the IMO and the propensity

[11] In fact, IMO participants from low-income score slightly higher and are somewhat more likely to win a medal. This result may seem counterintuitive as we expect training quality to be worse in developing countries but we note that participants from low-income countries might have higher incentives to perform well at the IMO. One should also keep in mind that many low-income countries do not send participants in the first place.

to a PhD in mathematics for participants from different groups of countries. In Figure 3, we plot the share of IMO participants obtaining a PhD in math by points scored at the IMO (five-points bands) across groups of countries. The general pattern we observe is that, for a given number of points, the share of participants getting a PhD in math is typically highest for high-income countries, followed by upper middle-income, then by lower middle-income, with low-income countries having the lowest share.

We investigate cross-country differences more formally using the following specification:

$$Y_{it} = \beta_1 IMOscore_{it} + \beta_2 CountryIncomeGroup_i + \eta_t + \varepsilon_{it} \tag{2}$$

where $i$ indexes medalists and $t$ indexes Olympiad years. $Y_{it}$ is an indicator variable for getting a PhD in mathematics, getting a PhD in mathematics from a top school, log(publications+1) and log(citations+1). Our variable of interest is the income group of the country that a participant represented at the Olympiad. We include indicator variables for low-income, lower middle-income and upper middle-income with high-income the omitted category. Crucially, we control for the number of points scored at the IMO, our proxy for talent.[12] We also control for Olympiad year fixed effects ($\eta_t$), focusing the comparison on individuals who participated in the same competition.

(insert Table 2 about here)

Results (see Table 2) suggest that across all long-term productivity outcome variables IMO participants from low- and middle-income countries significantly underperform compared to their high-income counterparts. For instance, IMO participants from low-income countries are 16 percentage points less likely to do a PhD and 3.2 percentage points less likely to do a PhD in a top

[12]We find similar results throughout if instead of controlling for the IMO score linearly we include fixed effects for every possible IMO score.

school; they produce 34% fewer publications and 56% fewer cites. To put things in perspective, IMO participants from low-income countries are only about half as likely to get a PhD from a top school when compared to rich country counterparts (with the same IMO score). A similar, though less pronounced, pattern can be observed for participants from middle-income countries. Considering the manually collected employment data on medalists, IMO participants from low-income and lower middle-income are (unsurprisingly in light of the previous results) less likely to be employed in mathematics academia (table 2 column 5). Moreover, we do not find evidence that IMO participants from low-income and lower middle-income are more likely to be employed in non-mathematics academia (column 6) or in visible industry positions (column 7).

The fact that developing country IMO participants are less likely to get a PhD in mathematics, and produce less mathematical knowledge, is compatible with multiple potential explanations. One possibility is that in poorer countries, mathematics competitions attract a pool of participants who are less interested in mathematics for its own sake and more in using success in competition to get access to universities or otherwise advance their careers. Alternatively, IMO participants from developing countries may train more extensively for the IMO than those from developed countries; in this case they may have lower mathematical ability conditional on achieving the same IMO scores. However, while IMO participants from developing countries may have higher incentives than those of developed countries to perform well at the IMO, the training infrastructure (including mathematics camps and material) is presumably better in richer countries. Finally, of course, differences in the environment could be driving the cross-country differences in mathematics careers and knowledge production. Poorer countries tend to have worse universities, which hampers learning and reduces the attractiveness of academic careers. Resources constraints may also lead individuals to choose more immediately lucrative careers outside mathematics and science.

How much knowledge production (from IMO participants) could there be if IMO participants from low-income countries were producing knowledge at the same rate as those from high-income countries? To estimate the size of the loss, we multiply the coefficients on the country income groups in our main specifications by the share of IMO participants in each group. We conclude that the knowledge production (of IMO participants) could be 10%[13] higher in terms of publications and 17% in terms of cites if IMO participants from low-income countries were producing knowledge at the same rate as IMO participants from high-income countries. While this calculation suggests that the benefits of enabling individuals from developing country to produce knowledge at the same rate as those from developed countries may be sizeable, we are unable to quantify the costs of doing so, and such costs may be substantial. On the one hand, a number of targeted fellowships for particularly talented individuals, or spots in highly ranked mathematical programs would not be particularly expensive. On the other hand, improving mathematical training and research in developing countries could involve larger costs.

# V Conclusion

This paper studies two questions about the role of talent in the advancement of the knowledge frontier. First, how does knowledge produced over a lifetime depend on talent displayed in teen years? Second, conditional on a given level of talent in teenage years, what is the impact of country of birth on knowledge produced? We focus on knowledge production in mathematics, as this allows us to use a unique institutional feature of the discipline – the IMO – to overcome empirical challenges to answering these two questions. By following IMO participants over their lifetime, we are able to measure talent in a comparable way across multiple countries and also to

[13]The share of IMO participants from low-income countries is 11% while 23% of participants are from lower-middle income countries and a further 18% from upper middle countries. Multiplying these shares by the coefficients from Table 2 column 3, we get: 0.11*(-0.337)+0.23*(-0.194)+0.18*(-0.083)=-0.096. The calculation for cites is analogous. The calculation for cites is analogous.

construct a sample without selecting on eventual lifetime success in knowledge production.

We document a strong and consistent link between IMO scores and a number of achievements in mathematics, including getting a PhD, mathematics publications and cites, and being awarded a Fields medal. That is, even in this group of teens who fall in the extreme right tail of the talent distribution, small differences in talent translates into sizeable differences in long-term achievements. We show that IMO participants from low- and middle-income countries produce consistently less mathematical knowledge than equally talented participants from high-income countries. Our results suggest that the quantity of lost knowledge production arising from cross-country differences in the productivity of IMO participants is sizeable, and that this lost knowledge production is not easily replaceable by that of other mathematicians.

While we have devoted considerable attention to the link between IMO scores and future performance in mathematics, we are in no way implying that high ability in problem solving in late teenage years – much less IMO participation or performance – is a necessary condition to become a successful mathematician. Some individuals may excel at mathematics knowledge production without scoring well on IMO-style tests or having a taste for that type of competition. Others may become interested in mathematics after their teenage years. We simply use the IMO medals as a tool to observe part of the extreme right tail of the ability distribution. What we are suggesting is the knowledge frontier could advance faster if individuals in the extreme right tail of the ability distribution – some of whom are IMO participants and some of whom are not – do not drop out of scientific careers.

One might question whether less mathematical knowledge produced is in any way bad for welfare. An in-depth analysis or even discussion of the contributions of mathematics to the economy is beyond the scope of this paper. However, we note that there is plenty of anecdotal evidence of mathematical discoveries having direct or indirect practical applications, such as in weather simulations, cryptography and telecommunications.

Even if mathematical knowledge production did contribute to welfare, the lost knowledge production arising from the under-utilization of developing-country talent is more palatable (or perhaps even desirable) if talent from developing countries is used to produce other types of knowledge. We cannot rule out the possibility that developing country talent ends up in valuable occupations (outside mathematical and non-mathematical knowledge production) where they might make distinctive contributions. However, if we think of IMO participants as having a strong natural comparative advantage in one very particular activity (mathematics) – as we do – then this makes it more likely that the current allocation is inefficient.

Having more developing country talent engaged into mathematics knowledge production might have side effects on the production of developed country mathematicians. On the one hand, there might be learning and spillovers (Azoulay et al. 2010). On the other hand, competition from developing country talent might induce displacement and crowding out of developed country mathematicians if there is a limited number of spots in graduate schools or faculty ranks. Borjas and Doran (2012) document such effects among American mathematicians following the influx of talented Russian mathematicians after the collapse of the Soviet Union. These effects may be important for the distribution of welfare among knowledge producers (and potential knowledge producers). From the perspective of advancing the knowledge frontier, however, it is highly desirable to have the most talented people engaged in knowledge production: as we show in this paper, they have a disproportionate ability to make ground-breaking contributions.

This paper falls short of identifying precisely why developing country participants are less likely to become professional mathematicians and produce less mathematical knowledge. Research and training capacity in the home country is likely to play a role. However, other factors may also be at play. For instance, developing country participants may have different preferences or private incentives to enter different types of careers, in particular if careers outside mathematics pay more. Future research may further elucidate the role of different factors in cross-country differences in the utilization of talent. For now, we briefly mention several types of push programs (or supply-side policies) that could be useful in light of the findings of the paper. First, fellowships for high-end talent to study mathematics at undergraduate and/or graduate levels may alleviate resources constraints and make mathematics careers more attractive. Second, top schools could encourage applications from developing countries; and recruiting elite talent to their student programs is probably in their interest. Third, strengthening mathematics research and training capacity in developing countries could not only improve the training of those who prefer to stay in their home country, but would also make mathematics research careers more attractive to them.

This paper has focused on the intensive margin of unutilized talent. We study individuals who reached excellence as teenagers and document differences in their subsequent career and knowledge production depending on the country they come from. However, there are also massive differences across countries in the number of individuals who reach excellence as teenagers.[14] High-income countries represent only around 15% of the world population but account for 48% of IMO participants and 64% of IMO gold medalists. Meanwhile, many geniuses from poor countries are never discovered or given the chance to excel as teenagers in the first place. This is an extensive margin of unutilized talent. Finding and supporting these geniuses is presumably difficult and could be costly, but also has great potential for the advancement of the knowledge frontier.

[14] Similarly large differences exist on the extensive margin with respect to gender. Even in the 2018 IMO, only 10% of participants were women. Ellison and Swanson (2010, 2018) study the gender gap in American mathematics competitions.

# References

Aghion, Philippe, Akcigit, Ufuk, Hyytinen, Ari, and Otto Toivanen. 2017. “The Social Origins of Inventors.” NBER Working Paper No 24110. https://doi.org/10.3386/w24110

Akcigit, Ufuk, Grigsby, John, and Tom Nicholas. 2017. “The Rise of American Ingenuity: Innovation and Inventors of the Golden Age.” NBER Working Paper No 23047. https://doi.org/10.3386/w23047

Agarwal, Ruchir, and Patrick Gaule. 2020. “Replication data for: Invisible Geniuses: Could the Knowledge Frontier Advance Faster?” *American Economic Association* [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E117424V1

Agrawal, Ajay, Goldfarb Avi, and Florenta Teodoridis. 2016 “Understanding the Changing Structure of Scientific Inquiry.” *American Economic Journal: Applied Economics* 8, no. 1: 100-128. http://dx.doi.org/10.1257/app.20140135

Andreescu, Titu, Gallian, Joseph A., Kane, Jonathan M., and Janet E Mertz. 2008. “Cross-cultural analysis of students with exceptional talent in mathematical problem solving.” *Notices of the AMS* 55, no. 10: 1248-1260. https://doi.org/doi=10.1.1.469.9929

Azoulay, Pierre, Graff Zivin, Josh S., and Jialan Wang. 2010. “Superstar extinction.” *Quarterly Journal of Economics* 125, no. 2: 549-589. https://doi.org/10.1162/qjec.2010.125.2.549

Azoulay, Pierre, Graff Zivin, Josh S., and Gustavo Manso. 2011 “Incentives and creativity: evidence from the academic life sciences.” *The RAND Journal of Economics* 42, no 3: 527-554. https://doi.org/10.1111/j.1756-2171.2011.00140.x

Bell, Alex, Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, and John Van Reenen. 2019 “Who Becomes an Inventor in America? The Importance of Exposure to Innovation” *Quarterly Journal of Economics* 134, no 2: 647-713. https://doi.org/10.1093/qje/qjy028

Baumol, William J. 1990. “Entrepreneurship: Productive, Unproductive, and Destructive.” *Journal of Political Economy* 98, no.5: 893-921. https://doi.org/10.1086/261712

Bischof, Daniel. 2017. “New Graphic Schemes for Stata: plotplain & plottig.”, *Stata Journal* 17, no. 3:748759. https://doi.org/10.1177/1536867X1701700313

Bloom, Nicholas, Jones, Charles I., Van Reenen, John, and Michael Webb (2017). “Are ideas getting harder to find?” NBER Working Paper No 23782. National Bureau of Economic Research. https://doi.org/10.3386/w23782

Borjas, George J., and Kirk B. Doran. 2012. “The Collapse of the Soviet Union and the Productivity of American Mathematicians.” *Quarterly Journal of Economics* 127, no. 3: 1143-1203. https://doi.org/10.1093/qje/qjs015

—. 2015a. “Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas.” *Journal of Labor Economics* 33, no S1: S109-S145. https://doi.org/10.1086/676659

—. 2015b “Which peers matter? The relative impacts of collaborators, colleagues, and competitors.” *Review of Economics and Statistics* 97, no 5: 1104-1117. https://doi.org/10.1162/REST_a_00472

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. “Robust nonparametric confidence intervals for regression-discontinuity designs.” *Econometrica* 82, no. 6: 2295-2326. https://doi.org/10.3982/ECTA11757

Celik, Murat Alp. 2017. “Does the Cream Always Rise to the Top? The Misallocation of Talent and Innovation.” Unpublished.

Cox, C. M. 1926. Genetic studies of genius. II. The early mental traits of three hundred geniuses. Stanford University Press

Ellison, Glenn, and Ashley Swanson. 2010. “The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24, no. 2: 10928. https://doi.org/10.1257/jep.24.2.109

—. 2016. "Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions." *American Economic Review* 106, no. 6: 1244-77. https://doi.org/10.1257/aer.20140308

—. 2018. "Dynamics of the Gender Gap in High Math Achievement." NBER Working Paper No 24910. https://doi.org/10.3386/w24910

Frandsen, Brigham R. 2017. "Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete." In Regression Discontinuity Designs: Theory and Applications (pp. 281-315). Emerald Publishing Limited.

Freeman, Richard, and John Van Reenen. 2009. "What if Congress doubled R&D spending on the physical sciences?" *Innovation Policy and the Economy*. 9, no 1: 1-38. Chicago: University of Chicago Press. https://doi.org/10.1086/592419

Fu, Qiang, Lu, Jingfeng, and Lu Yuanzhu . 2012. "Incentivizing R&D: Prize or subsidies?" *International Journal of Industrial Organization* 30, no 1: 67-79. https://doi.org/10.1016/j.ijindorg.2011.05.005

Ganguli, Ina. 2017. "Saving Soviet science: The impact of grants when government R&D funding disappears". *American Economic Journal: Applied Economics* 9, no. 2: 165-201. https://doi.org/10.1257/app.20160180

Hsieh, Chang-Tai, Hurst, Erik, Jones, Charles I., and Peter J. Klenow. 2019. "The allocation of talent and U.S. economic growth." *Econometrica* 87, no. 5: 1439-1474. https://doi.org/10.3982/ECTA11427

Iaria, Alessandro, Schwarz, Carlo, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science." *Quarterly Journal of*

*Economics* 133, no. 2: 927-991. https://doi.org/10.1093/qje/qjx046

Kremer, Michael, and Heidi Williams. 2010. “Incentivizing innovation: Adding to the toolkit.” *Innovation Policy and the Economy* 10, no. 1: 1-17. Chicago: University of Chicago Press. https://doi.org/10.1086/605851

Jacob, Brian A., and Lars Lefgren. 2011. “The impact of research grant funding on scientific productivity.” *Journal of Public Economics* 95, no. 9: 1168-1177. https://doi.org/10.1016/j.jpubeco.2011.05.005

Kahn, Shulamit, and Megan J MacGarvie. 2016. “How Important is US Location for Research in Science?” *Review of Economics and Statistics* 98, no. 2: 397-414. https://doi.org/10.1162/REST_a_00490

Jones, Charles I. 2002. “Sources of US Economic Growth in a World of Ideas.” *American Economic Review* 92, no. 1: 220-239. https://doi.org/10.1257/000282802760015685

Maurer, Stephen M. 2006. “Choosing the right incentive strategy for research and development in neglected diseases.” *Bulletin of the World Health Organization* 84, no. 5: 376-381. https://doi.org/10.2471/BLT.06.029835

Murphy, Kevin M., Shleifer, Andrei, and Robert W. Vishny. 1991. “The Allocation of Talent: Implications for Growth.” *Quarterly Journal of Economics* 106, no. 2: 503-530. https://doi.org/10.2307/2937945

Stephan, Paula E. 2012. “How Economics Shapes Science”. Cambridge, MA: Harvard University Press.

Ugarov, Alexander. 2019. “Talent Misallocation across Countries: Evidence from Educational Achievement Tests.” Unpublished manuscript.

Waldinger, Fabian. 2010. “Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany.” *Journal of Political Economy* 118, no. 4: 787-831.

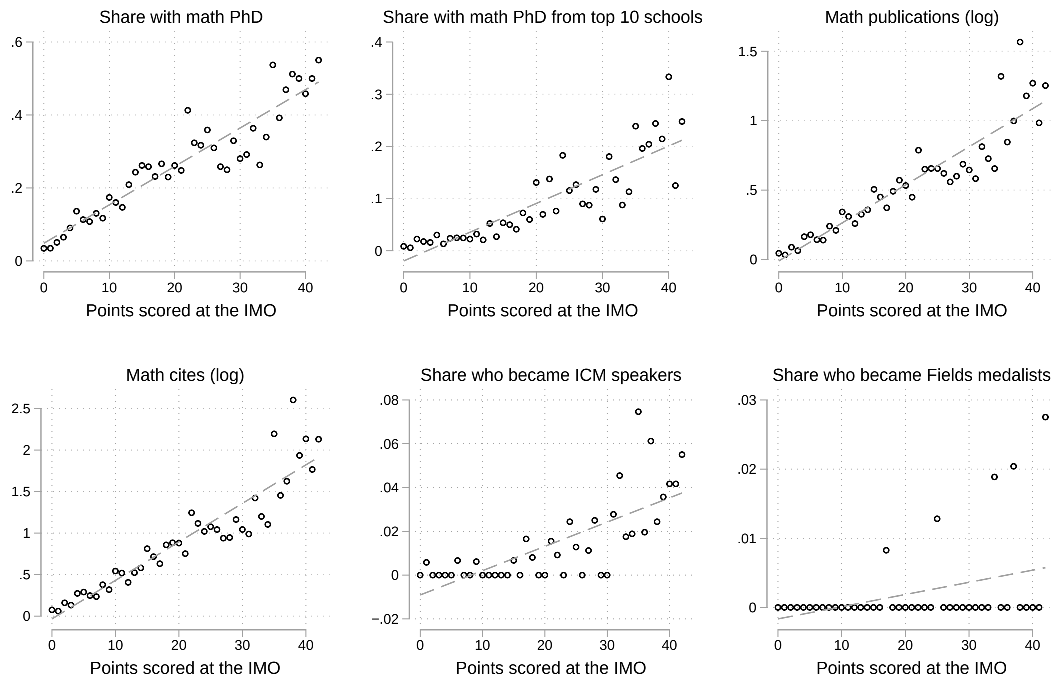
https://doi.org/10.1086/655976

—. 2011. “Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany.” *Review of Economic Studies* 79, no. 2: 838-861. https://doi.org/10.1093/restud/rdr029

Williams, Heidi. 2012. “Innovation inducement prizes: Connecting research to policy.” *Journal of Policy Analysis and Management* 31, no. 3: 752-776. https://doi.org/10.1002/pam.21638

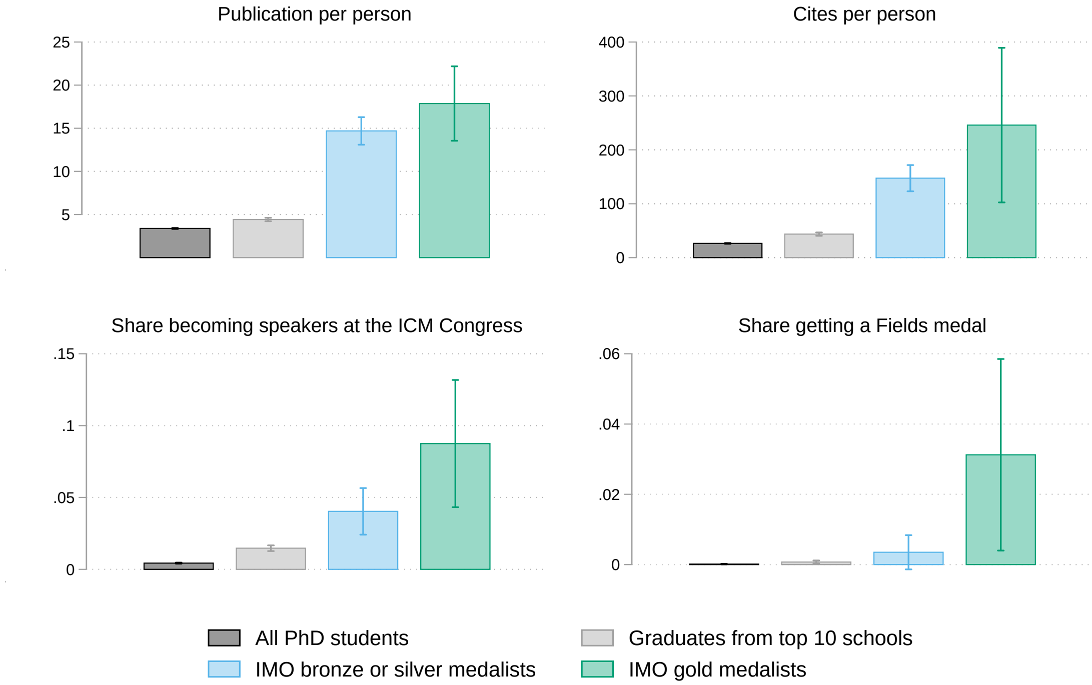
# Exhibits

**Figure 1:** Relationship between points scored at the IMO and subsequent achievements




*Notes*: We compute the sample means of each of the six outcomes variables by the number of points scored at the IMO (for publications and cites, we add one before taking the logs, and then compute the sample means). We then plot the resulting number against the number of points scored. A linear fit is superimposed.

**Figure 2:** Comparing IMO medalists with other professional mathematicians



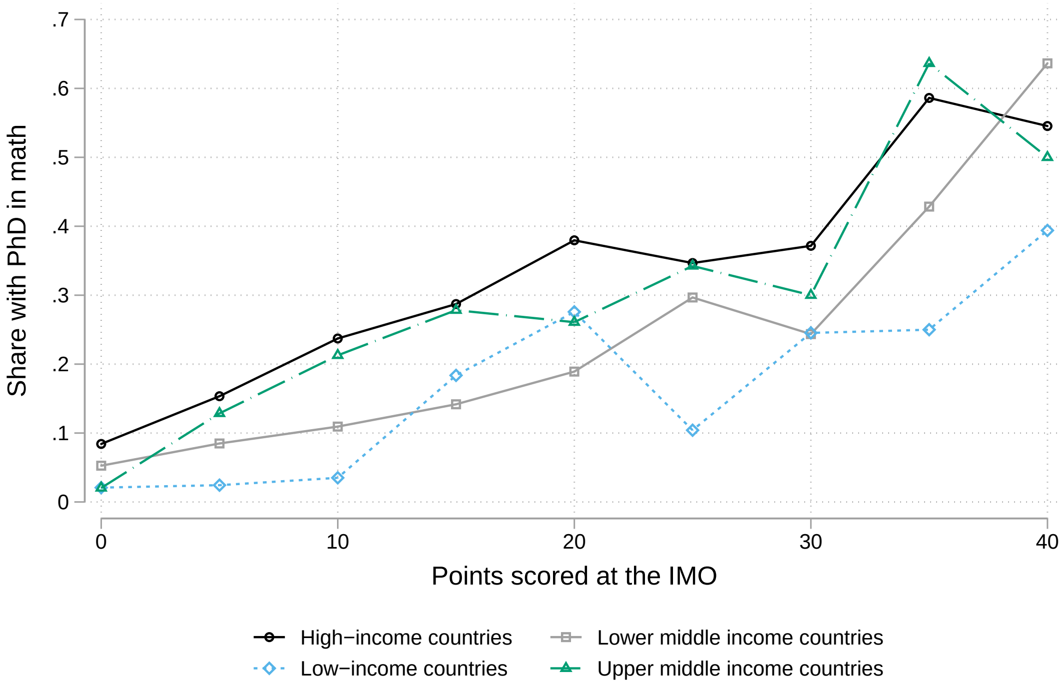

*Notes*: These figures are based on an ancillary sample including all mathematics PhD graduates listed in the Mathematics Genealogy Project. The graphs display sample means across four outcomes (publications, cites, becoming a speaker at the ICM congress, and becoming a Fields medalist) for four groups of PhD graduates (all PhD graduates, PhD graduates from top ten schools, IMO bronze or silver medalists, IMO gold medalists).

**Table 1:** IMO score and medal receipt by country income group

| | (1) IMO Score | (2) Received any medal |
|---|---|---|
| Income group (country of origin): | | |
| Low-income | 3.206 | 0.093 |
| | (0.583) | (0.024) |
| Lower middle-income | 0.623 | 0.006 |
| | (0.403) | (0.019) |
| Upper middle-income | 0.030 | 0.022 |
| | (0.430) | (0.021) |
| High-income: omitted | | |
| Observations | 4,710 | 4,710 |
| Mean of D.V. for high-income participants | 15.841 | 0.482 |

*Notes*: Both regressions are estimated by OLS and include cohort (olympiad year) fixed effects. The income level groups based on the World Bank 2000 classification and the omitted country income level category is high-income countries. Robust standard errors in parentheses.

**Figure 3:** Share getting a PhD in mathematics across country income groups




*Notes*: We compute the share of IMO participants getting a PhD in mathematics by the number of IMO points scored (5-points bands) and plot the resulting share against the number of points scored.

**Table 2:** Link between IMO score, long-term performance, and occupation by country income group

| | (1) Math PhD | (2) Math PhD (top 10) | (3) Pubs | (4) Cites | (5) Academia (math) | (6) Academia (non-math) | (7) Industry |
|---|---|---|---|---|---|---|---|
| Income group (country of origin): | | | | | | | |
| Low-income | -0.152 | -0.031 | -0.337 | -0.560 | -0.111 | -0.011 | 0.039 |
| | (0.017) | (0.012) | (0.041) | (0.067) | (0.026) | (0.021) | (0.027) |
| Lower middle-income | -0.101 | -0.022 | -0.194 | -0.321 | -0.049 | -0.004 | -0.043 |
| | (0.014) | (0.009) | (0.034) | (0.057) | (0.023) | (0.017) | (0.021) |
| Upper middle-income | -0.040 | -0.025 | -0.083 | -0.171 | 0.002 | 0.082 | -0.064 |
| | (0.017) | (0.010) | (0.041) | (0.066) | (0.026) | (0.022) | (0.021) |
| High-income: omitted | | | | | | | |
| IMO Score | 0.011 | 0.005 | 0.027 | 0.045 | 0.007 | 0.002 | -0.001 |
| | (0.001) | (0.000) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Observations | 4710 | 4710 | 4710 | 4710 | 2272 | 2272 | 2272 |
| Mean of D.V. for high-income participants | 0.263 | 0.080 | 0.531 | 0.886 | 0.261 | 0.108 | 0.186 |

*Notes*: The dependent variables are as follows: obtaining a math PhD (column 1), obtaining a math PhD from a top 10 school (column 2), the log of mathematics publications plus one (column 3), the log of mathematics cites plus one (column 4), whether the person is currently employed in math academia (column 5), currently employed in non-math academia (column 6) or in industry with an online profile (column 7). All regressions are estimated by OLS and include cohort (olympiad year) fixed effects. The sample for column 1 to 4 is all IMO participants who competed between 1981 and 2000. For columns 5, 6 and 7 the sample is all IMO participants who competed between 1981 and 2000 and won a medal (the employment outcomes were only collected for medalists). The variables of interest are the income group levels of a participant's country of origin. The income level groups based on the World Bank 2000 classification and the omitted country income level category is high-income countries. Robust standard errors in parentheses.